

We claim:

1. A video processing device, comprising:
an audio event detecting means for detecting audio events in a video data;
and
a memory communicating with said audio event detecting means and storing video data and audio data corresponding to said video data;
wherein said audio event detecting means detects an audio event in said audio data and indexes said video data at about a beginning of said audio event.
2. The device of claim 1, wherein said processor indexes said video data by extracting and storing one or more representative video frames.
3. The device of claim 1, wherein said processor indexes said video data by inserting index data into said video data.
4. The device of claim 1, wherein said processor indexes said video data by saving one or more index pointers.
5. The device of claim 1, wherein said processor indexes said video data by recording one or more time stamps.
6. The device of claim 1, wherein said audio event comprises speech.
7. The device of claim 1, wherein said audio event comprises music.

8. The device of claim 1, wherein said video processing device comprises a video recorder device.

9. The device of claim 1, wherein said video processing device comprises a video editor device.

10. The device of claim 1, wherein said video processing device comprises a video authoring device.

11. A video processing device, comprising:
a processor;
an audio event detector communicating with said processor; and
a memory communicating with said processor, said memory storing video data and audio data corresponding to said video data;
wherein said audio event detector detects an audio event in said audio data and wherein said processor indexes said video data at about a beginning of said audio event.

12. The device of claim 11, wherein said video processing device comprises a video recorder device.

13. The device of claim 11, wherein said video processing device comprises a video editor device.

14. The device of claim 11, wherein said video processing device comprises a video authoring device.

15. The device of claim 11, wherein said processor indexes said video data by extracting and storing one or more representative video frames.

16. The device of claim 11, wherein said processor indexes said video data by inserting index data into said video data.

19. The device of claim 11, wherein said memory stores a predetermined energy threshold, a predetermined ZCR variance threshold, a predetermined ZCR amplitude span threshold, and a predetermined set of speech harmonics thresholds, and wherein said audio event detector further comprises:

an energy detector communicating with said processor and measuring an energy content of said audio data;

a ZCR detector communicating with said processor and generating a ZCR value from said audio data;

a spectrum detector communicating with said processor and generating a frequency spectrum from said audio data;

wherein said audio event detector compares harmonic frequency components in said frequency spectrum to said predetermined set of speech harmonics thresholds and detects a speech audio event if said harmonic frequency components fall within said predetermined set of speech harmonics thresholds if a ZCR value span exceeds said predetermined ZCR amplitude span threshold, if a variance between said ZCR value and one or more previous ZCR values is above the predetermined ZCR variance threshold, and if said energy content is greater than said predetermined energy threshold.

20. The device of claim 19, wherein said spectrum detector comprises an FFT processor.

21. The device of claim 11, wherein said memory stores a predetermined energy threshold and a predetermined frequency change threshold and wherein said audio event detector further comprises:

an energy detector communicating with said processor and measuring an energy content of said audio data;

a spectrum detector communicating with said processor and generating a frequency spectrum from said audio data;

a peak detector communicating with said processor and said spectrum detector, said peak detector receiving said frequency spectrum, detecting frequency peaks in said frequency spectrum, and generating a frequency peak output; and

wherein said audio event detector compares frequency peaks in two or more frequency peak outputs and detects a music audio event if said frequency peaks in said two or more frequency peak outputs are substantially stable and if said energy content is greater than said predetermined energy threshold.

22. The device of claim 21, wherein said spectrum detector comprises an FFT processor.

23. The device of claim 21, wherein said music detector compares frequency peaks in two or more consecutive frequency peak outputs.

24. A method of indexing a video data, comprising the steps of:
 detecting an audio event in an audio data corresponding to said video data;
 and
 indexing one or more representative video frames of said video data at about
 a beginning of said audio event.

25. The method of claim 24, with the step of detecting said audio event
 further comprising detecting a speech audio event in said audio data.

26. The method of claim 24, with the step of detecting said audio event
 further comprising the steps of:
 comparing an energy content of said audio data to a predetermined energy
 threshold;
 comparing a ZCR variance and a ZCR value span of said audio data to a
 predetermined ZCR variance threshold and to a predetermined ZCR amplitude span
 threshold, respectively, if said energy content is greater than said predetermined
 energy threshold;
 comparing harmonic frequency components of said audio data to a
 predetermined set of speech harmonics thresholds if said ZCR variance and said
 ZCR value span exceed said predetermined ZCR variance threshold and said
 predetermined ZCR amplitude span threshold, respectively; and
 detecting a speech audio event if said harmonic frequency components are
 within said predetermined speech harmonics range.

27. The method of claim 24, with the step of detecting said audio event further comprising detecting a music audio event in said audio data.

28. The method of claim 24, with the step of detecting said audio event further comprising the steps of:

comparing an energy content of said audio data to a predetermined energy threshold;

comparing frequency peaks in two or more consecutive frequency spectra if said energy content is greater than said predetermined energy threshold; and

detecting a music audio event if said frequency peaks in said two or more consecutive frequency spectra are substantially stable.

29. The method of claim 24, with the step of indexing further comprising indexing said video data at about a beginning of a semantically meaningful video scene.

30. The method of claim 24, with the step of indexing further comprising extracting and storing said one or more representative video frames.

31. The method of claim 24, with the step of indexing further comprising inserting index data into said video data.

32. The method of claim 24, with the step of indexing further comprising saving one or more index pointers.

33. The method of claim 24, with the step of indexing further comprising storing one or more time stamps.

33. The method of claim 24, with the step of indexing further comprising storing one or more time stamps.